Published in final edited form as:

J Cross Cult Psychol. 2019 July ; 50(6): 806–824. doi:10.1177/0022022119852422.

Cross-Country Differences in Parental Reporting of Symptoms of ADHD

Beatriz MacDonald,

University of New Mexico Health Sciences Center

Bruce F. Pennington,

University of Denver

Erik G. Willcutt,

University of Colorado Boulder

Julia Dmitrieva,

University of Denver

Stefan Samuelsson,

Linköping University and Stavanger University

Brian Byrne, and

University of New England and Linköping University

Richard K. Olson

University of Colorado Boulder and Linköping University

Abstract

Previous studies within the United States suggest there are cultural and contextual influences on how Attention-Deficit/Hyperactivity Disorder (ADHD) symptoms are perceived. If such influences operate within a single country, they are likely to also occur between countries. In the current study, we tested whether country differences in mean ADHD scores also reflect cultural and contextual differences, as opposed to actual etiological differences. The sample for the present study included 974 participants from four countries tested at two-time points, the end of preschool and the end of 2nd grade. Consistent with previous research, we found lower mean ADHD scores in Norway and Sweden in comparison to Australia and the United States, and we tested four explanations for these country differences: 1) Genuine etiological differences, 2) Slower introduction to formal academic skills in Norway and Sweden than in the United States and Australia that indicated a context difference, 3) Under-reporting tendency in Norway and Sweden, or 4) Over-reporting tendency in the United States and Australia. Either under-or over-reporting would be examples of cultural differences in the perception of ADHD symptoms. Of these explanations, results of ADHD measurement equivalence tests across countries rejected the first three explanations and supported the fourth explanation: an over-reporting tendency in the United States and Australia. These findings indicate that parental reporting of ADHD symptoms is more

accurate in Norway and Sweden than in Australia and the United States, and thus have important clinical and educational implications for how parental reporting informs an ADHD diagnosis in these countries.

Keywords

ADHD; Assessment; Cultural Considerations

Any behaviorally-defined disorder requires a somewhat arbitrary diagnostic cutoff on a continuous symptom dimension for its diagnosis. Hence, a cutoff that is valid in one culture might not be valid in another. In addition, even with the same cutoff, symptom endorsements could vary by culture, since patients, parents, and teachers depend on their own internal threshold for what counts as a problematic behavior. Hence, while the underlying liability for such disorders may be a valid neurodevelopmental dimension across countries, country differences in diagnostic cutoffs and/or internal thresholds for symptom endorsements across countries inevitably involve cultural and contextual conventions. Thus, biology and culture are unavoidably intertwined in the definition and diagnosis of behaviorally-defined disorders. This fact makes cross-cultural and cross-country findings of differences in rates and/or symptoms of such disorders difficult to interpret. Do such differences mean that the countries and cultures in question differ in the actual underlying liability for such disorders, in the perception of them, or in some combination of both (such as a gene by environment interaction)? The present study addresses this broad question in the specific context of mean symptom levels of Attention-Deficit/Hyperactivity Disorder (ADHD) across countries.

The worldwide-pooled prevalence estimate of ADHD is 5.29%, but rates among countries differ (Polanczyk, deLima, Horta, Biederman, & Rohde, 2007). The percentage of school-age children affected by ADHD is reportedly lower in Sweden (3.7%; Kadesjo & Gillberg, 2001) and Norway (5.2%; Ullebo, Breivik, Gillberg, Lundervold, & Posserud, 2012) in comparison to the United States (7–9%; Akinbami, Liu, Pastor, & Reuben, 2011; Willcutt et al., 2012) and Australia (6–7%; Graetz, Sawyer, Hazell, Arney, & Baghurst, 2001). Differences in prevalence estimates of ADHD are typically explained by methodological differences, such as diagnostic criteria, source of information, requirement of impairment for diagnosis, and geographic location within countries (Polanczyk et al., 2007; Polanczyk, Willcutt, Salum, Kieling, & Rohde, 2014). Such methodological explanations make sense when study design is confounded with country, as is true in meta-analyses across countries. If country differences are still found when the same methods are used in a single study of different countries, then other explanations are called for, such as an actual etiological difference between countries, a cultural difference in the perception of symptoms, or some combination of the two explanations.

Previous research using a broad measure of emotional and behavioral problems has consistently found country differences *within* the same study (Crijnen, Achenbach, & Verhulst, 1999; Rescorla et al., 2007; Verhulst et al., 2003), suggesting there are possibly etiological or perceptual differences across countries. A valid country difference in ADHD rates means that the balance of etiological risk and protective factors differs on average

between countries (e.g. socioeconomic status and genetics). The etiology of a disorder refers to its original causes in a population of individuals, that is the genetic and environmental risk and protective factors that alter brain and behavior development to produce the disorder. Non-valid group differences, therefore, are not etiological but instead due to sample or measurement artifacts. Rescorla and colleagues (2007) compared parents' ratings of behavioral and emotional problems on the Child Behavior Checklist (CBCL) for general population samples of children ages 6 to 16 from 31 societies ($N = 55,508$). Overall, Sweden, Norway, Germany, Iceland and Japan reported fewer problems than other societies, such as the United States and Australia. Norway and Sweden scored similarly and more than 1 standard deviation (SD) better than the overall mean across countries of total problems scores, whereas the United States and Australia scored worse than this overall mean. The reported means were as followed for each country: Sweden ($M = 14.35$, $SD = 13.10$), Norway ($M = 15.31$, $SD = 12.85$), Australia ($M = 18.38$, $SD = 18.79$) and the United States ($M = 23.65$, $SD = 18.81$). Interestingly, these differences applied to ADHD symptoms on the CBCL but were not specific to them, suggesting either a general etiological or perceptual difference across countries, or some combination of the two explanations.

These studies considered but did not test explanations for the observed country differences. For instance, countries may vary in their expectations for how children should behave, resulting in the same behavior being rated as problematic in one country but not in another (Rescorla et al., 2007). In this case, the country difference is due to cultural perception. Another possible explanation is based on the range of socio-economic status (SES) in each country. Countries that have a relatively small range of SES differences, such as Norway and Sweden, may have lower behavioral and emotional symptom levels than countries that have larger ranges, such as the United States (Achenbach, 2010), because of the adverse effects of low SES on child behavioral development. This would be an example of an etiological difference between countries. What is novel about the current study is that it tested whether country differences in reported ADHD symptoms are due to etiological differences or differences in cultural perception (or some combination) by parents. Moreover, the current study used a more specific measure of parental ratings of ADHD symptoms (i.e. the Disruptive Behaviors Rating Scale; DBRS) versus the CBCL, which is a broadband measure of externalizing and internalizing behaviors.

Four Possible Explanations

Four possible explanations of the ADHD behavioral rating difference between countries are presented as ovals in the decision tree in Figure 1, along with the empirical tests that distinguish them. We could reject a fifth possible explanation for the country differences in ADHD ratings, namely a selection artifact. Participants for the current study were recruited from community settings in each country and in the same fashion across countries. In addition, similar results from many other community samples across countries indicate that the country differences in mean symptom levels of ADHD are not due to a selection artifact, especially since the difference persists regardless of the behavioral rating measure used (Crijnen et al., 1999; Faraone, Sergeant, Gillberg, & Biederman, 2003; Heiervang, Goodman, & Goodman, 2008; Polanczyk et al., 2007; Polanczyk et al., 2014; Rescorla et al., 2007; Verhulst et al., 2003).

The first question in the decision tree in Figure 1, is whether the country difference persists across time points. The context hypothesis, which was proposed by Willcutt and colleagues (2007), posited that the early educational context (i.e., preschool and kindergarten) makes ADHD-related behaviors more salient in some countries than others, but once the educational context becomes similar in later grades, the country difference in salience is eliminated. Parents and teachers in Norway and Sweden are not as involved in structured literacy activities in preschool and kindergarten as is the case in the United States and Australia (S. Samuelsson, personal communication, 2010). Thus, attentional difficulties might not be as apparent in preschool in Norway and Sweden in comparison to Australia and the United States. If this context hypothesis is true, then the mean symptom levels of ADHD would not differ once children in Norway and Sweden begin structured literacy activities and schooling. In the current study, we tested this context hypothesis by comparing ADHD behavioral ratings collected both at the end of preschool and in 2nd grade. The context hypothesis predicted a grade by country interaction in the mean symptom levels of ADHD, such that the country differences found in preschool would diminish significantly by 2nd grade.

If the context hypothesis could be rejected, the next test in the decision tree in Figure 1 was a test of measurement equivalence across countries. This is the crucial test for distinguishing between a non-artifact difference versus a perceptual explanation of the country differences. Procedures to evaluate measurement equivalence have been extensively used to evaluate possible test bias (Vandenberg & Lance, 2000) across ethnic or cultural groups, but the same procedures can be usefully applied to examine measurement equivalence across countries. Procedures include internal and external reliability analyses, such as variance range, internal consistency, test-retest stability, inter-rater reliability, confirmatory factor analysis, analysis of individual items, and external validity.

If there is measurement equivalence across countries, then the genuine etiologic differences hypothesis would be supported, essentially by elimination. Providing positive evidence for an etiologic difference across countries with higher levels of ADHD symptoms would require a different study design than what is used in the present study and would involve directly testing putative etiological risk factors.

If there is not measurement equivalence across countries, the two remaining hypotheses involve cultural differences in sensitivity to ADHD symptoms. This is to say that raters in different countries differ in their sensitivity to these symptoms because they may perceive the same behaviors through different cultural *lenses*. For instance, some cultures may be more reluctant than others to report “real” behavioral and emotional problems, hence an under-reporting tendency in those cultures. In contrast some cultures may pathologize typical behaviors, hence an over-reporting tendency in those cultures. In this case, the stigma associated with mental illness, especially in children (Hinshaw, 2005; Kieling et al., 2011), might be greater in the under-reporting countries. Distinguishing these two hypotheses requires examining the external validity of ADHD ratings in each country. Either tendency, over-or under-reporting, should reduce the external validity of ADHD symptom ratings relative to cultures without such a tendency. Hence, if the external validity is greater in countries with lower mean symptom levels of ADHD compared to those with higher mean

levels, then that would indicate an over-reporting tendency in the countries with higher mean levels. In contrast, if the external validity is greater in countries with higher mean levels, then that would indicate an under-reporting tendency in the countries with lower mean levels. In this study, the external validator of ADHD behavioral ratings was actual performance on cognitive measures known to be sensitive to ADHD, specifically rapid naming, verbal memory, and visuospatial skill (Doyle, Biederman, Seidman, Weber, & Faraone, 2000; Mealer, Morgan, & Luscomb, 1996). To examine the outlined hypotheses, we used a longitudinal sample (International Longitudinal Twins Study, ILTS) of twins that is part of a larger international project; for our purposes we selected specific time points (end of preschool and 2nd grade) (see Byrne et al., 2002; Byrne et al., 2006; Byrne et al., 2007; Christopher et al., 2013; Olson et al., 2011; Samuelsson et al., 2005).

Method

Sample

The current study used a population-based sample of preschool age twins from the United States, Australia, Norway and Sweden, who were followed into early school age (see Byrne et al., 2002; Byrne et al., 2006; Byrne et al., 2007; Christopher et al., 2013; Olson et al., 2011; Samuelsson et al., 2005). The International Longitudinal Twin Sample (ILTS) included 974 twin pairs from the National Health and Medical Research Council's Australian Twin Registry, the Medical Birth Registries in Norway and Sweden, and the Colorado Twin Registry in the United States. The families in Australia, Norway and Sweden were all approached by mail and had a participation rate of 60%. Families of the Colorado Twin Registry were approached by mail or phone and 86% agreed to participate. Only participants whose first language was the native language of their country were selected. At the end of preschool, Australia had recruited 253 twin pairs, the United States had 482 twin pairs, and Norway and Sweden had 239 twin pairs. At the time of analysis, some participants had not completed some of the 2nd grade measures. At the end of 2nd grade, Australia had 94 twin pairs, the United States had 375 twin pairs, and Norway and Sweden had 128 twin pairs. Since the focus of the current study was parental reporting and not genetic etiology, the sample was limited to one participant selected at random from each twin pair to simulate a population sample and correct for familiarity.

At initial contact and testing, all twins were in their final preschool year, with ages ranging from 47 to 68 months ($M = 57.8$) in Australia, from 58 to 68 months ($M = 60.1$) in Norway and Sweden, and from 54 to 71 months ($M = 58.8$) in the United States. There were no differences in gender distribution across countries (Australia: 139 males & 125 females; United States: 243 males & 246 females; Norway and Sweden: 121 males & 127 females). None of the children were enrolled in kindergarten at the first time point, but a majority of them were attending a preschool program (94% in Australia, 90% in Norway and Sweden, and 85% in the United States). Literacy activities were excluded from the national preschool curriculum of Norway and Sweden. In contrast, most of the preschool programs in Australia and the United States included at least some literacy activities, although the specific activities varied across programs. By 2nd grade, teaching environment and literacy practices were more similar across the four countries (Byrne et al., 2007).

There were no significant differences across countries for parental mean years of education in Australia ($M = 13.4$), Norway and Sweden ($M = 13.9$), and the United States ($M = 14.1$). However, it must be noted that only 10% of twins are currently registered in Australia. On national socioeconomic status (SES) indicators with a mean of 100 and standard deviation of 15, the Australian Twins Registry scores are: for education, $M = 101.4$ ($SD = 11.9$); for income, $M = 100.5$ ($SD = 11.6$); for occupation, $M = 101.5$ ($SD = 7.5$); and for single parenthood, $M = 100.6$ ($SD = 6.1$). These numbers indicate a slight bias toward higher SES and less variance for families in Australia; however, the difference is small and non-significant. Additionally, the sample from Norway and Sweden has a high percentage of participants from rural areas.

Measures

Translation of tests.—All testing materials in the study were originally in English. Approximately ninety percent of the material in the tests was translated without having to change the original words or pictures. The Disruptive Behavior Rating Scale (DBRS) had already been translated to Norwegian and Swedish but the translations had not been formally published in their respective countries.

For the cognitive measures, the serial naming tasks and subtests derived from the Wide Range Assessment of Memory and Learning (WRAML; Adams & Sheslow, 1990) required no or only limited work on translations. However, tests measuring nonword repetition and morphology knowledge were all translated. Morphological rules are very similar across languages (Samuelsson et al., 2005), and similar morphemes were used for the Norwegian and Swedish versions of the tests. Therefore, the high degree of overlap increased the comparability between Norwegian and Swedish and English measures. The versions from the subtests of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI -Revised; Wechsler, 1989) had been previously translated and standardized for both Norway and Sweden.

Behavioral rating scale.—The Disruptive Behavior Rating Scale (DBRS; Barkley & Murphy, 1998) was used to obtain parent and teacher ratings of the 18 symptoms of ADHD according to the DSMS-IV-TR (American Psychiatric Association, 2000). The scale has two dimensions with 9 symptoms to assess “inattention” and 9 symptoms to assess “hyperactivity and impulsivity”. Each symptom is rated on a four-point scale (0-*never or rarely*, 1-*sometimes*, 2-*often*, and 3-*very often*). Items rated as *often* or *very often* were scored as positive symptoms and items rated as *never or rarely* or *sometimes* were scored as negative symptoms. The DBRS and other similar ADHD rating scales (ADHD-RS-IV) are reliable and valid measures used in routine clinical care across Europe, North America, and Australia (Dopfner et al., 2008; Dopfner et al., 2006; Faraone, Doyle, Mick, & Biederman, 2001; Faraone et al., 2003; Ford, Goodman, & Meltzer, 2003; Marzocchi et al., 2004) and in Latin American countries (Rohde et al., 2001).

For the current study, we used the parental endorsement on the DBRS to determine the mean symptom levels of ADHD, which is consistent with previous cross-country studies that use parental report (Crijnen et al., 1999; Rescorla et al., 2007; Verhulst et al., 2003; Willcutt et

al., 2007). Further, a diagnosis of ADHD made based on parental report alone, on average, positively predicts a diagnosis based on teacher report (Biederman, Faraone, Milberger, & Doyle, 1993; Biederman, Keenan, & Faraone, 1990; Zeiner, 1997). Previous results from other studies indicate that parent and teacher ratings on the DBRS or other similar ADHD scales are internally consistent [e.g. for teacher and parent ratings, $\alpha = .92-.96$, (Pelham, Gnagy, Greenslade, & Milich, 1992); for parent ratings, $\alpha = .86-.91$, (Willcutt et al., 2007)]. However, concordance between teacher and parental reporting is found to be relatively poor with low rater agreement (Gomez, 2007; Mitsis, McKay, Schulz, Newcorn, & Halperin, 2000) in different countries. Further, in the research literature, “case” families of ADHD are identified based on parent and/or teacher symptom ratings (Willcutt, 2012; Willcutt, Pennington, Olson, Chhabildas, & Hulslander, 2005). The current study defined the High ADHD group as participants who at the end of preschool had a severity score on the DBRS of 25 points or higher on both dimensions (inattention and hyperactivity/impulsivity) based on parental report, and the Low ADHD group had a score below 25 points. In contrast, for a proper clinical diagnosis of ADHD, it is recommended for the assessment to include behavioral ratings, a clinical semi-structured interview, and possible objective measures of sustained attention. Therefore, we defined our High ADHD group as a high-risk group.

Cognitive variables.—To test the external validity of ADHD in each country, we chose the following measures of Verbal Memory, Naming Speed, and Visuospatial Skill, administered at the end of preschool. For the purposes of this study, a brief description of each measure is provided; a complete description of each measure is available in previous papers (Byrne et al., 2002; Samuelsson et al., 2005). Composite measures of each construct were calculated based on prior factor analyses and theoretical considerations (Samuelsson et al., 2005). Scores on each measure were standardized across the entire sample while controlling age so that we were able to assess mean differences on the cognitive variables across countries. Each composite score was then calculated by computing the mean of the child’s standardized scores on the tests included in the composite.

Verbal memory.—There were three measures of this construct: the Nonword Repetition task developed by Gathercole, Willis, Baddeley and Emslie (1994) (Cronbach’s $\alpha = .84$), the Sentence Memory subtest from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) battery (Wechsler, 1989) (split-half coefficient reliability coefficient of .88 for 5-year-old children), and the Sound Symbol subtest from the Wide Range Assessment of Memory and Learning (WRAML; Adams & Sheslow, 1990) (Cronbach’s $\alpha = .88$).

Naming speed.—The Rapid Object Naming and Rapid Color Naming subtests from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) were used as two measures of naming speed (Cronbach’s $\alpha = .81$ for color naming and .71 for object naming).

Visuospatial skill.—This construct was measured by the Block Design subtest from the WPPSI (Wechsler, 1989) (published split-half coefficient is .86 for this subtest) and the Visual Learning subtest from the Wide Range Assessment of Memory and Learning (WRAML; Cronbach’s $\alpha = .85$) (Adams & Sheslow, 1990).

Results

Data Cleaning and Preliminary Analyses

One twin from each pair was selected at random to control for non-independence. All scores were examined for outliers, and outliers were winsorized to 3 *SD*. The raw ADHD scores were used for further analysis, whereas the cognitive measures were corrected for age. Possible linear and nonlinear effects of age on the cognitive measures were controlled by regressing the raw scores of each variable on to age and age squared and saving the standardized residuals. The residuals were standardized across the entire sample in order to maintain country differences.

At both time points, the Norwegian and Swedish samples performed similarly on the DBRS measures and cognitive variables without significant statistical differences. Therefore, in the following analyses, the samples of Norway and Sweden were combined and treated as one sample labeled Scandinavia. The theoretical rationale for combining Norway and Sweden is their very similar language, as well as educational and social policies at the time of testing participants. On this basis, we were able to combine the samples to maximize statistical power (see previous published articles using Scandinavia as one sample, Byrne et al., 2002; Byrne et al., 2006; Byrne et al., 2007; Christopher et al., 2013; Olson et al., 2011; Samuelsson et al., 2005).

Ten percent of the United States sample ($n = 48$), eleven percent of the Australian sample ($n = 26$), and four percent of the Scandinavian sample ($n = 10$) were classified as having a severity score of 25 or higher on the DBRS at the end of preschool. These percentages are similar to the prevalence rates of clinically diagnosed cases of ADHD in each of the countries (Akinbami et al., 2011; Graetz et al., 2001; Kadesjo & Gillberg, 2001; Ullebo et al., 2012; Willcutt, 2012).

Group Demographics

There were differences in participant age at end of preschool, with Scandinavian participants being on average a few months older than children in the United States and Australia ($F[2, 997] = 133.76, p < .001$; Australia: age $M = 57.2$ months, $SD = 3.4$; United States: $M = 58.8$, $SD = 2.3$; Scandinavia: $M = 60.9$, $SD = 1.7$). Therefore, analyses modeled age as an independent variable to control for age effects. As expected, ADHD symptom severity was significantly higher in males than females for both time points. At the end of preschool, males had more severe scores than females on total ADHD severity in Australia (male $M(SD) = 14.58(8.29)$; female $M(SD) = 12.49(7.74)$; $t(251) = 2.07, p < .05$), in the United States (male $M(SD) = 14.36(8.72)$; female $M(SD) = 11.54(7.56)$; $t(480) = 3.89, p < .001$), and in Scandinavia (male $M(SD) = 11.67(7.05)$; female $M(SD) = 9.04(6.97)$; $t(237) = 2.89, p < .01$). At the end of 2nd grade, males also had more severe scores than females on total ADHD severity in Australia (male $M(SD) = 12.28(8.11)$; female $M(SD) = 8.95(5.47)$; $t(93) = 2.34, p < .05$), in the United States (male $M(SD) = 11.31(8.69)$; female $M(SD) = 7.75(6.40)$; $t(373) = 4.54, p < .001$), and in Scandinavia (male $M(SD) = 8.55(6.99)$; female $M(SD) = 5.71(6.39)$; $t(126) = 2.40, p < .05$).

Main Analyses

The four hypotheses were tested in a sequential order, using SPSS – Version 19 if not otherwise specified. Testing the context hypothesis (Hypothesis 1) required examining the stability of country differences across both time points. Testing the remaining three hypotheses required examining measurement invariance across countries.

Context hypothesis (Hypothesis 1).—An initial set of analyses were conducted to ensure that we replicated the previous preschool results of the study of Willcutt et al. (2007), which only examined the preschool age group in a smaller sample. Using parental report, mean symptom levels of ADHD were significantly higher in Australia and the United States than in Scandinavia at the end of preschool (see Table 1), consistent with the results of Willcutt et al. (2007).

Our new finding was that this country difference was also present at the end of 2nd grade, and there was no country by time interaction effect (see Table 1). This finding did not support the context hypothesis that predicted that raters for all three countries would report similar levels of ADHD behaviors once the literacy practices were more comparable at the end of 2nd grade.

Measurement equivalence.—The tests of hypothesis 2 (genuine etiologic differences), hypothesis 3 (under-reporting tendency) and hypothesis 4 (over-reporting tendency) required evaluating the measurement equivalence of the ADHD rating scale (DBRS) across countries, in terms of both external and internal validity.

External validity.—Analyses included a mixed model ANOVA with age, country and ADHD group (High versus Low ADHD status, severity rating higher than or equal to 25 versus lower than 25) as between group factors, and each of the three cognitive composites (Verbal Learning, Naming Speed, Visuospatial Skill) as the within-group dependent variables. The results (see Table 2) were similar across composites in finding main effects of country ($F[2, 961] = 20.01, p < .001$; $F[2, 963] = 4.87, p < .01$; and $F[2, 965] = 14.36, p < .001$, respectively) and High versus Low ADHD status ($F[1, 961] = 15.13, p < .001$; $F[1, 963] = 12.91, p < .001$; and $F[1, 965] = 4.70, p < .05$, respectively); there was no main effect of age. As can be seen in Table 2, the main effect for country was due to Australia consistently having the highest scores across composites, and Scandinavia consistently having the lowest. The main effect of ADHD status was due to the High ADHD groups consistently performing worse than the Low ADHD groups across countries, a finding which validates each of these cognitive composites as a cognitive correlate of ADHD. Finally, the ADHD status main effect was moderated by country, either significantly ($F[2, 961] = 4.09, p < .05$) for Verbal Memory or as a trend ($F[2, 963] = 2.08, p < .10$) for Naming Speed. This interaction effect was due to the ADHD main effect having the largest effect size for Verbal Memory and Naming Speed in Scandinavia (Cohen's $d = 1.12$ and 0.77 , respectively), whereas these effect sizes were considerably lower in the United States and Australia (d values ranging from 0.06 to 0.41 , see Table 2). For Visuospatial Skill, the United States had the largest effect size (Cohen's $d = .47$) in comparison to the effect sizes in Scandinavia and Australia (Cohen's $d = .36$ and $.07$, respectively).

Comparing the effect sizes for the High ADHD group main effect across countries in Table 2 provided support for Hypothesis 4 (over-reporting tendency hypothesis). These effect sizes were considerably larger in Scandinavia ($d = .36 - 1.12$) than in the other two countries ($d = .06 - .47$), whose effect sizes were similar to each other. To confirm this apparent difference in external validity, the correlations between cognitive composites and DBRS severity ratings were compared across countries. As expected from the differences in effect sizes, Scandinavia had higher external validity correlations across all three cognitive composites (Scandinavia: Verbal Memory $r = -.24, p < .001$, Naming Speed $r = -.22, p < .001$, Visuospatial Skill $r = -.16, p < .05$; United State: Verbal Memory $r = -.12, p < .05$, Naming Speed $r = -.08, p < .15$, Visuospatial Skill $r = -.12, p < .01$; Australia: Verbal Memory $r = -.06, n.s.$, Naming Speed $r = -.16, p < .01$, Visuospatial Skill $r = -.05, n.s.$). To test whether the difference between correlation coefficients across countries was significant, we converted each correlation into a z -score using Fisher's r -to- z transformation. Then, making use of the sample size employed to obtain each coefficient, the z -scores were compared. Of all nine comparisons, the comparison between Australia and Scandinavia for the Verbal Memory composite variable was significantly different, and the comparison between Scandinavia and the United States for the Naming Speed variable was a trend. These results suggested stronger external validity of ADHD ratings in Scandinavia, such that raters in Scandinavia reported behavioral problems more accurately, consistent with Hypothesis 4, an over-reporting tendency in Australia and the United States.

This over-reporting tendency could affect any one of the internal validators. For instance, differential reporting in one country could not only alter the variance in the measure (and alter its kurtosis) in comparison to the other countries, but also affect the other three tests of internal validity. Hence, we next examined measurement equivalence across countries with respect to internal validity.

Internal validity.—Tests of internal validity included 1) kurtosis and skewness, 2) reliability, 3) factor structure, and 4) rank order of endorsement levels for individual items. For reliability analyses (test-rest stability and inter-rater reliability), data only included participants that had ADHD ratings for both time points or both raters.

Kurtosis and skewness. Using raw ADHD severity rating scores from parents, all skewness values were within an acceptable range for each country and the variances were similar across countries (Levene's Tests of Equality of Error Variances). These findings also do not support Hypothesis 3, an under-reporting tendency in Scandinavia, since such a tendency restricts the variance of the distribution and increases kurtosis.

Internal consistency. The internal consistency of the DBRS was calculated for symptom dimensions and total severity ratings per parental report. All Cronbach's alphas were satisfactory for each country (values ranging from .84 to .92) and for both time points (end of preschool and end of 2nd grade). Scandinavia was not consistently the lowest, again not supporting Hypothesis 3 (under-reporting tendency). Since the internal consistency is similar across countries, measurement equivalence was supported by this indicator.

Test-retest stability.: Intra-class correlations (ICCs) for Australia, the United States, and Scandinavia indicated significant levels of stability for all ADHD variables (ICCs values ranging from .30 to .62). Although the test-retest correlations were somewhat higher in Scandinavia than the United States, there were no significant differences, again supporting measurement equivalence.

Inter-rater reliability.: Inter-rater reliability for the subtype and total severity ratings of the DBRS were computed. The raters were parents and teachers and ratings were collected for both time points. The ICCs ranged from .13 to .32 at the end of preschool, and .25 to .58 at the end of 2nd grade. All ICCs were statistically significant. Although Scandinavia had higher correlations, there was not a significant difference among countries. Although the ICCs appeared to be low for rater agreement, they were consistent with previous literature that compared ratings of teachers and parents using the DBRS (Gomez, 2007).

In sum, these four reliability analyses found no internal validity differences across the three countries. In particular, no support was found for Hypothesis 3 (under-reporting tendency). Nonetheless, raters in these three countries could differ on how they use different individual items on the DBRS and that issue was examined next.

Confirmatory factor analysis.: Factor analysis was used to test construct equivalence of the DBRS across the three countries. The 18-item ADHD set was subjected to confirmatory factor analysis (CFA) to determine whether the two-factor (inattention and hyperactivity/impulsivity) solution for ADHD items reported by Willcutt and colleagues (2012) and others would be confirmed in these samples. In addition, a multi-group analysis of invariance was conducted to test the factor structure and loadings across Australia, Scandinavia, and the United States. The steps followed to test measurement invariance included: configural invariance, weak factorial (metric) invariance, and comparison of loading stability between two time points (end of preschool and end of 2nd grade). Parental report of the DBRS was used for this set of analysis. AMOS (Version 16) was used to perform the CFAs. To correct for missing data across time points (attrition), full information maximum likelihood estimation was selected for each model.

Table 3 shows the means, standard deviations, and loadings for each DBRS item for the two separate time points (i.e., end of preschool and end of 2nd grade). At the end of preschool for all three countries, the one-factor model had a poor fit: $\chi^2(135) = 1849.8, p < .001$; CFI = .760; RMSEA = .114 (see Table 4). Adding the second factor resulted in a significant improvement to the model fit: $\chi^2(1) = 869.7, p < .001$. However, the two-factor model still did not have adequate fit: $\chi^2(134) = 980.1, p < .001$; CFI = .882; RMSEA = .081. Based on the modification indices, five pairs of correlations were added among the items of the same factor – items 1 and 2, items 7 and 9, items 11 and 12, items 15 and 16, and items 17 and 18. The final two-factor model had a good fit: $\chi^2(129) = 633.3, p < .001$; CFI = .929; RMSEA = .063. All variables had significant factor loadings (see Table 3). We examined invariance of this two-factor model across the three countries. The two-factor model had a good fit in each country (See Table 4 for model fit statistics), indicating configural invariance.

Next, we examined the invariance of factor loadings across the three countries (Table 5) at the end of preschool. The model with factor loadings constrained for all countries had a significantly poorer fit than the unconstrained model: $\chi^2(32) = 70.02, p > .001$. Consequently, three models were tested to assess which countries differed. The model that constrained item loadings for the United States and Australia was not significantly different than the unconstrained model ($\chi^2[16] = 25.11, p = .07$). However, the two models that constrained item loadings for Scandinavia with item loadings for Australia or the United States were significantly different than the unconstrained model (Australia and Scandinavia constrained: $\chi^2[16] = 33.81, p < .01$; United States and Scandinavia constrained: $\chi^2([16] = 46.46, p < .001$). These results suggested that the factor loadings are similar for the United States and Australia, but different in Scandinavia, indicating that the United States and Australia exhibited weak factorial invariance, but Scandinavia did not. Similar results were found at the end of 2nd grade. This result means the ADHD measure is being used differently by raters in Scandinavia compared to the other two countries.

In sum, results from the CFAs indicated that raters in Scandinavia, Australia, and the United States separated symptoms of ADHD into two dimensions: inattention and hyperactivity/impulsivity, indicating configural invariance. This was true for all three countries at the end of preschool and at the end of 2nd grade. However, the ADHD rating scale demonstrated weak factorial invariance only across Australia and the United States. In contrast, ADHD symptoms loaded differently for Scandinavia. This was the first internal indication that raters in Scandinavia are using the ADHD measure differently than raters in the other two countries, consistent with a country difference in how ADHD symptoms are perceived.

Stability of factor loadings of the DSM–IV ADHD symptoms.: We also examined the stability of factor loading for each country separately (Table 6). We again found similar country differences, with the United States and Australia demonstrating similar factor loadings for attention across time points, but dissimilar factor loadings for hyperactivity/impulsivity. Interestingly, Scandinavia's models exhibited the opposite pattern. The hyperactivity/impulsivity factor loadings were not significantly different across time, but the inattention factor loadings differed significantly between end of preschool and end of 2nd grade.

Analysis of individual items.: We next conducted a rank-order correlation analysis of rates of item endorsement, which ranks correlations from highest to lowest for each country for comparison to the rankings of the other two countries. The order varied across countries but some of the top ranked items were similar for all the countries (See Table 3 for means and standard deviations). However, rank order correlations between the United States and Australia were stronger than those with Scandinavia. Rank order correlations for the end of preschool ranged from .11 to .97 with the highest correlation between United States and Australia and the weakest correlation between Australia and Scandinavia. All three countries ranked the inattention items, “easily distracted” and “doesn’t listen,” and hyperactive/impulsive items, “difficulty waiting turn” and “interrupts,” within the top endorsed items. For the end of 2nd grade, the rank order correlations were more consistent ranging from .41

to .68. All three countries ranked the inattention items, “easily distracted” and “fails to attend,” and hyperactive/impulsive items, “difficulty waiting turn” and “interrupts.”

In sum, the internal validity analyses indicated a lack of measurement equivalence between Scandinavia and the other two countries, Australia and the United States. Hence, raters in Scandinavia used the ADHD rating scale somewhat differently than raters in the other two countries. These internal validity results converged with external validity results in indicating a lack of measurement equivalence between Scandinavia and the other two countries, which did exhibit measurement equivalence. Since there was not measurement equivalence across countries, we rejected the genuine etiologic differences hypothesis (Hypothesis 2). We had already rejected the context hypothesis (Hypothesis 1) because the country differences in ADHD severity ratings persisted until the end of 2nd grade. Given that the ADHD behavioral ratings had greater external validity in Scandinavia than in Australia and the United States, the over-reporting tendency hypothesis (Hypothesis 4) was supported and the under-reporting tendency hypothesis (Hypothesis 3) was rejected.

Discussion

The current study examined four competing hypotheses to explain the lower mean symptom levels of ADHD in Norway and Sweden in comparison to Australia and the United States. These hypotheses were 1) a context difference, 2) a genuine etiologic difference, 3) an under-reporting tendency in Norway and Sweden, or 4) an over-reporting tendency in Australia and the United States. The first two hypotheses were readily rejected. Contrary to the context hypothesis suggested by Willcutt et al. (2007), the country differences in mean symptoms level of ADHD were stable across time points, even as the school context in Norway and Sweden changed. Norway and Sweden performed the worst of the four countries on the three cognitive composites, despite having the best (lowest) ADHD severity ratings. Genuine etiologic differences required that Norway and Sweden perform the best on *both* ADHD severity ratings and cognitive correlates.

Evaluating the remaining over-and under-reporting tendencies hypotheses required testing for measurement equivalence of the ADHD rating scale (DBRS) across the three countries using tests of both internal and external validity. We found there was not complete measurement equivalence across countries. Overall, the United States and Australia had similar constructs of ADHD based on the DBRS. Symptoms of inattention exhibited stability across the two time points, but certain symptoms of hyperactivity/impulsivity were stronger indicators of the construct at the end of 2nd grade compared to the end of preschool. One interpretation of this pattern is that symptoms of inattention were equally salient to parents in these two countries at both time points, but some symptoms of hyperactivity/impulsivity became more salient (were viewed as more deviant) at the end of 2nd grade. Scandinavia, on the other hand, appeared to have a different construct of ADHD in which certain symptoms of inattention, not of hyperactivity/impulsivity, became more salient across time points. These differences in Norway and Sweden could reflect different behavioral norms for behavior at school and at home than in the other two countries in this time period. These results, again, could be consistent with either over-or under-reporting of symptoms of ADHD.

Although the variance and reliabilities of the ADHD rating scale (DBRS) were similar across countries, the factor loadings and rank order of item endorsements were different in Norway and Sweden compared to that of the other two countries. Most importantly, the ADHD rating scale (DBRS) had stronger external validity in Scandinavia, which supported the over-reporting tendency hypotheses versus the under-reporting tendency hypothesis.

The finding of an over-reporting tendency in Australia and the United States as compared to Norway and Sweden is a novel finding. The finding converges with results from studies examining differences in ADHD rates as either a function of age at school entry (e.g. Elder, 2011; Morrow et al., 2012) or education policy differences among states in the United States (Fulton et al., 2009; Nigg, 2006). Although previous studies have found lower mean symptoms levels of ADHD in Scandinavian countries, they have not tested the explanation for those lower ratings, as was done in the present study.

The findings of the current study have clinical implications for the United States and other countries. At a public health level, we need to re-assess the threshold being operationalized to diagnose children with ADHD. We recommend that these findings be incorporated in training programs for mental health professionals and medical practitioners so that they can be more informed about the cross-cultural differences in symptom reporting. It is possible that ADHD is being over-diagnosed in some countries, such as the United States and Australia. The study of Angold, Erkanli, Egger and Costello (2000) found that the rate of stimulant treatment in the United States was twice the rate of parent-reported ADHD, based on a standardized psychiatric interview. Interestingly, many children who did meet criteria for ADHD were not being treated and the majority of children and adolescents who were receiving stimulants did not fully meet criteria for an ADHD diagnosis.

The current study may need to be replicated using the new DSM-5 criteria to diagnose ADHD, which has been described as having a more lenient threshold in certain age groups. It is imperative to know if the DSM-5 changes will have an impact in the use of parental reporting of ADHD symptoms, which is a necessary component in making a clinical diagnosis. Mental health stigma may also influence the over-or under-reporting tendency of behavioral and emotional symptoms between countries. People's denial of mental health problems and their reluctance to access services is often the result of stigma. For instance, lower behavioral and emotional questionnaire scores in children from Scandinavian countries have been speculated to reflect the stigma of mental illness (Hinshaw, 2005; Kielsing et al., 2011). Parents are often reluctant for mental health providers to use labels to describe problematic behaviors (Corrigan, 2004). In the National Stigma Study for Children, a significant percentage of adults in the United States who correctly identified the vignettes of children either displaying ADHD or depression symptoms readily rejected the mental health illness label (Pescosolido et al., 2008). However, adults in the United States were less likely to view ADHD as a serious mental health illness or as a diagnosis with stigma, which may explain the over-reporting tendency identified in the current study.

Possible limitations of the current study include sample recruitment because the team in the Scandinavian countries recruited mostly from rural areas and the Australian sample had a slight bias toward higher SES and less variance for families. However, other population

studies have found similar reporting tendencies and mean symptoms levels (Crijnen et al., 1999; Rescorla et al., 2007; Verhulst et al., 2003). When working with translated materials, translation effects are always a limitation. However, considering that other studies (Crijnen et al., 1999; Rescorla et al., 2007; Verhulst et al., 2003) have found the same effect, it is overly pessimistic to assume that all measurement translations of English to Norwegian or Swedish are incorrect. Finally, the age range of the sample of the study was of younger children (preschool to 2nd grade), which might limit the findings. However, the same effect was found in a population sample of 6 to 16-year-olds (Crijnen et al., 1999; Rescorla et al., 2007; Verhulst et al., 2003). Therefore, future research should examine the over-reporting tendency that has been identified.

Acknowledgments

Funding was provided by the National Institutes of Health (P50 HD027802 for the Colorado Learning Disabilities Research Center and R01 HD038526 for the Colorado component of the International Longitudinal Twin Study [ILTS]). The Australian component of the ILTS was facilitated through access to the Australian Twin Registry, a national resource supported by an Enabling Grant (628911) from the National Health and Medical Research Council. Funding was provided by the Australian Research Council (DP0663498 and DP0770805). The Scandinavian component of the ILTS was supported by the Research Council of Norway (154715/330), the Swedish Research Council (345–2002-3701, PDOKJ028/2006:1, and 2011–1905), and the Swedish Council for Working Life and Social Research (2011–0177). We thank the twins and their families who participated in our research.

References

- Achenbach TM (2010). Multicultural evidence-based assessment of child and adolescent psychopathology. *Transcultural Psychiatry*, 47, 707–726. doi:10.1177/1363461510382590 [PubMed: 21088101]
- Adams W, & Sheslow D (1990). *Wide range assessment of memory and learning*. Wilmington, DE: Jastak Associates.
- Akinbami LJ, Liu X, Pastor PN, & Reuben CA (2011). Attention deficit hyperactivity disorder among children aged 5–17 years in the United States, 1998–2009 NCHS Data Brief (70, Centers for Disease Control and Prevention), 1–8.
- Angold A, Erkanli A, Egger HL, & Costello EJ (2000). Stimulant treatment for children: a community perspective. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 975–984; discussion 984–994. doi:10.1097/00004583-200008000-00009 [PubMed: 10939226]
- Association AP (2000). *Diagnostic and statistical manual of mental disorders (4th-TR ed.)*. Washington D.C: American Psychiatric Association.
- Barkley RA, & Murphy KR (1998). *Attention-deficit hyperactivity disorder: A clinical workbook (2nd ed)*. New York, NY: Guilford Press.
- Biederman J, Faraone SV, Milberger S, & Doyle A (1993). Diagnoses of Attention-deficit Hyperactivity Disorder from Parent R Predict Diagnoses Based on Teacher Reports. *Journal of the American Academy of Child & Adolescent Psychiatry*, 32, 315–317. [PubMed: 8444759]
- Biederman J, Keenan K, & Faraone SV (1990). Parent-based diagnosis of attention deficit disorder predicts a diagnosis based on teacher report. *Journal of the American Academy of Child & Adolescent Psychiatry*, 29, 698–701. [PubMed: 2228921]
- Byrne B, Delaland C, Fielding-Barnsley R, Quain P, Samuelsson S, Høien T, ... Olson RK (2002). Longitudinal twin study of early reading development in three countries: Preliminary results. *Annals of Dyslexia*, 52, 47–73.
- Byrne B, Olson RK, Samuelsson S, Wadsworth S, Corley R, DeFries JC, & Willcutt E (2006). Genetic and environmental influences on early literacy. *Journal of Research in Reading*, 29, 33–49.

- Byrne B, Samuelsson S, Wadsworth S, Hulslander J, Corley R, DeFries JC, ... Olson RK (2007). Longitudinal twin study of early literacy development: Preschool through Grade 1. *Reading and Writing: An Interdisciplinary Journal*, 20, 77–102.
- Christopher ME, Hulslander J, Byrne B, Samuelsson S, Keenan JM, Pennington BF, ... Olson RK (2013). The genetic and environmental etiologies of individual differences in early reading growth in Australia, the United States, and Scandinavia. *Journal of Experimental Child Psychology*, 115, 453–467. [PubMed: 23665180]
- Corrigan P (2004). How stigma interferes with mental health care. *American psychologist*, 59, 614. [PubMed: 15491256]
- Crijnen AA, Achenbach TM, & Verhulst FC (1999). Problems reported by parents of children in multiple cultures: the Child Behavior Checklist syndrome constructs. *American Journal of Psychiatry*, 156, 569–574. [PubMed: 10200736]
- Dopfner M, Breuer D, Wille N, Erhart M, Ravens-Sieberer U, & group B. s. (2008). How often do children meet ICD-10/DSM-IV criteria of attention deficit-/hyperactivity disorder and hyperkinetic disorder? Parent-based prevalence rates in a national sample--results of the BELLA study. *European Child and Adolescence Psychiatry*, 17 Suppl 1, 59–70. doi:10.1007/s00787-008-1007-y
- Dopfner M, Steinhausen HC, Coghill D, Dalsgaard S, Poole L, Ralston SJ, ... Group AS (2006). Cross-cultural reliability and validity of ADHD assessed by the ADHD Rating Scale in a pan-European study. *European Child and Adolescent Psychiatry*, 15 Suppl 1, I46–55. doi:10.1007/s00787-006-1007-8 [PubMed: 17177016]
- Doyle AE, Biederman J, Seidman LJ, Weber W, & Faraone SV (2000). Diagnostic efficiency of neuropsychological test scores for discriminating boys with and without attention deficit–hyperactivity disorder. *Journal of consulting and clinical psychology*, 68, 477. [PubMed: 10883564]
- Elder TE (2011). The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates. *Journal of Health Economics*, 29, 641–656.
- Faraone SV, Doyle AE, Mick E, & Biederman J (2001). Meta-analysis of the association between the 7-repeat allele of the dopamine D4 receptor gene and attention deficit hyperactivity disorder. *American Journal of Psychiatry*, 158, 1052–1057. [PubMed: 11431226]
- Faraone SV, Sergeant J, Gillberg C, & Biederman J (2003). The worldwide prevalence of ADHD: Is it an American condition? *World Psychiatry*, 2, 104–113.
- Ford T, Goodman R, & Meltzer H (2003). The British child and adolescent mental health survey 1999: The prevalence of DSM-IV disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 1203–1211.
- Fulton H, Scheffler R, Hinshaw S, Levine P, Stone S, Brown T, & Modrek S (2009). National variation of ADHD diagnostic prevalence and medication use: Health care providers and education policies. *Psychiatric Services*, 60, 1075–1083. [PubMed: 19648195]
- Gathercole SE, Willis CS, Baddeley AD, & Emslie H (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, 2, 103–127. [PubMed: 7584287]
- Gomez R (2007). Australian parent and teacher ratings of DSM-IV ADHD symptoms: Differential symptom functioning and parent-teacher agreement and differences. *Journal of Attention Disorders*, 11, 17–27. [PubMed: 17606769]
- Graetz BW, Sawyer MG, Hazell PL, Arney F, & Baghurst P (2001). Validity of DSM-IV ADHD subtypes in a nationally representative sample of Australian children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1410–1417. doi: 10.1097/00004583-200112000-00011
- Heiervang E, Goodman A, & Goodman R (2008). The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. *Journal of Child Psychology and Psychiatry*, 49, 678–685. doi:10.1111/j.1469-7610.2008.01882.x [PubMed: 18489678]
- Hinshaw SP (2005). The stigmatization of mental illness in children and parents: Developmental issues, family concerns, and research needs. *Journal of Child Psychology and Psychiatry*, 46, 714–734. [PubMed: 15972067]

- Kadesjo B, & Gillberg C (2001). The comorbidity of ADHD in the general population of Swedish school-age children. *Journal of Child Psychology and Psychiatry*, 42, 487–492. [PubMed: 11383964]
- Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, ... Rahman A (2011). Child and adolescent mental health worldwide: evidence for action. *Lancet*, 378, 1515–1525. doi: 10.1016/S0140-6736(11)60827-1 [PubMed: 22008427]
- Marzocchi GM, Capron C, DiPietro M, Tauleria ED, Duyme M, & Frigerio A (2004). The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. *European Child and Adolescent Psychiatry*, 13, ii40–ii46. [PubMed: 15243785]
- Mealer C, Morgan S, & Luscomb R (1996). Cognitive functioning of ADHD and non-ADHD boys on the WISC-III and WRAML: An analysis within a memory model. *Journal of Attention Disorders*, 1, 133–145.
- Mitsis EM, McKay KE, Schulz KP, Newcorn JH, & Halperin JM (2000). Parent-teacher concordance for DSM-IV attention-deficit/hyperactivity disorder in a clinic-referred sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39, 308–313. doi: 10.1097/00004583-200003000-00012 [PubMed: 10714050]
- Morrow RL, Garland EJ, Wright JM, Maclure M, Taylor S, & Dormuth CR (2012). Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *Canadian Medical Association Journal*, 184, 755–762. [PubMed: 22392937]
- Nigg JT (2006). *What causes ADHD?: Understanding what goes wrong and why*. New York: Guilford Press.
- Olson RK, Keenan JM, Byrne B, Samuelsson S, Coventry WL, Corley R, ... Hulslander J (2011). Genetic and environmental influences on vocabulary and reading development. *Scientific Studies of Reading*, 15, 26–46. [PubMed: 21132077]
- Pelham WE, Gnagy EM, Greenslade KE, & Milich R (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 31, 210–218. [PubMed: 1564021]
- Pescosolido BA, Jensen PS, Martin JK, Perry BL, Olafsdottir S, & Fettes D (2008). Public knowledge and assessment of child mental health problems: findings from the National Stigma Study-Children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47, 339–349. [PubMed: 18216729]
- Polanczyk GV, deLima MS, Horta BL, Biederman J, & Rohde LA (2007). The worldwide prevalence of ADHD: a systematic review and a meta-regression analysis. *The American Journal of Psychiatry*, 164, 942–948. [PubMed: 17541055]
- Polanczyk GV, Willcutt EG, Salum GA, Kieling C, & Rohde LA (2014). ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *International Journal of Epidemiology*, 43, 434–442. [PubMed: 24464188]
- Rescorla L, Achenbach T, Ivanova MY, Dumenci L, Almqvist F, Bilenberg N, ... Verhulst F (2007). Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *Journal of Emotional and Behavioral Disorders*, 15, 130–142. doi: 10.1177/10634266070150030101
- Rohde LA, Barbosa G, Polanczyk G, Eizirik M, Rasmussen ER, Neuman RJ, & Todd RD (2001). Factor and latent class analysis of DSM-IVADHD symptoms in a school sample of Brazilian adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 711–718. doi:10.1097/00004583-200106000-00017
- Samuelsson S, Byrne B, Quain P, Wadsworth S, Corley R, DeFries JC, ... Olson R (2005). Environmental and genetic Influences on prereading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology*, 97, 705–722.
- Ullebo AK, Breivik K, Gillberg C, Lundervold AJ, & Posserud MB (2012). The factor structure of ADHD in a general population of primary school children. *Journal of Child Psychology and Psychiatry*, 53, 927–936. doi:10.1111/j.1469-7610.2012.02549.x [PubMed: 22512532]
- Vandenberg RJ, & Lance CE (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70.

- Verhulst FC, Achenbach TM, van der Ende J, Erol N, Lambert MC, Leung PW, ... Zubrick SR (2003). Comparisons of problems reported by youths from seven countries. *American Journal of Psychiatry*, 160, 1479–1485. [PubMed: 12900311]
- Wagner RK, Torgesen JK, & Rashotte CA (1999). CTOPP: Comprehensive Test of Phonological Processing. Austin, TX: PRO-ED.
- Wechsler D (1989). Wechsler Preschool and Primary Scale of Intelligence--Revised. New York: Psychological Corporation.
- Willcutt EG (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics*, 9, 490–499. [PubMed: 22976615]
- Willcutt EG, Betjemann RS, Wadsworth SJ, Samuelsson S, Corley R, DeFries J, ... Olson RK (2007). Preschool twin study of the relation between attention deficit/hyperactivity disorder and prereading skills. *Reading and Writing: An Interdisciplinary Journal*, 20, 103–125.
- Willcutt EG, Nigg JT, Pennington BF, Solanto MV, Rohde LA, Tannock R, ... Lahey BB (2012). Validity of DSM-IV attention deficit/hyperactivity disorder Symptom dimensions and subtypes. *Journal of Abnormal Psychology*, 121, 991–1010. [PubMed: 22612200]
- Willcutt EG, Pennington BF, Olson RK, Chhabildas N, & Hulslander J (2005). Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: In search of the common deficit. *Developmental neuropsychology*, 27, 35–78. [PubMed: 15737942]
- Zeiner P (1997). Parent-reported symptoms of hyperactivity and attention deficits predict teacher-reported symptoms. *Acta Paediatrica*, 86, 178–182. [PubMed: 9055889]

Figure 1.
Decision Tree for Testing the Four Hypotheses

Table 1

Mean scores on parent ratings of ADHD behaviors in the three samples

ADHD Measure	ADHD scores			Effect sizes between countries		
	End of Preschool			Australia- -US	Australia- -Scandinavia	Scandinavia- -US
	Australia <i>n</i> = 253 <i>M</i> (<i>SD</i>) (Range)	USA <i>n</i> = 482 <i>M</i> (<i>SD</i>) (Range)	Scandinavia <i>n</i> = 239 <i>M</i> (<i>SD</i>) (Range)			
Inattention	6.0 (4.3) (0–19)	6.0 (4.4) (0–19)	4.9 (3.8) (0–17)	.00	.27 **	.26 **
Hyperactivity/ Impulsivity	7.6 (4.9) (0–22)	7.0 (4.9) (0–22)	5.4 (4.3) (0–22)	0.12	.48 **	.34 **
Total ADHD	13.6 (8.1) (0–38)	12.9 (8.3) (0–38)	10.3 (7.1) (0–38)	0.09	.45 **	.34 **
End of 2 nd Grade						
	<i>n</i> = 95	<i>n</i> = 375	<i>n</i> = 128			
Inattention	5.8 (4.2) (0–19)	5.3 (4.5) (0–19)	4.2 (3.9) (0–16)	.11	.40 **	.25 **
Hyperactivity/ Impulsivity	4.7 (3.9) (0–17)	4.1 (4.1) (0–17)	3.0 (3.7) (0–17)	.15	.45 **	.28 **
Total ADHD	10.5 (7.0) (0–33)	9.4 (7.7) (0–33)	7.2 (6.8) (0–33)	.15	.48 **	.29 **

Note. Means and *SD* are based on raw ADHD scores** Significant difference between countries, $p < .01$

Table 2

Mean scores of External Validators by country and ADHD status at the End of Preschool

Status	Composite Score									
	Verbal Memory			Naming Speed			Visuospatial Skill			
	Australia (<i>n</i> = 253) <i>M</i> (<i>SD</i>)	USA (<i>n</i> = 476) <i>M</i> (<i>SD</i>)	Scandinavia (<i>n</i> = 238) <i>M</i> (<i>SD</i>)	Australia (<i>n</i> = 251) <i>M</i> (<i>SD</i>)	USA (<i>n</i> = 479) <i>M</i> (<i>SD</i>)	Scandinavia (<i>n</i> = 239) <i>M</i> (<i>SD</i>)	Australia (<i>n</i> = 251) <i>M</i> (<i>SD</i>)	USA (<i>n</i> = 479) <i>M</i> (<i>SD</i>)	Scandinavia (<i>n</i> = 239) <i>M</i> (<i>SD</i>)	
Low ADHD	.40 (.99)	-.15 (.92)	.01 (.98)	.07 (.97)	.07 (1.01)	-.12 (.93)	.37 (1.01)	-.09 (.93)	-.13 (1.01)	
High ADHD	.34 (.96)	-.52 (.89)	-1.10 (1.0)	-.25 (.89)	-.14 (1.11)	-1.05 (1.44)	.30 (1.03)	-.53 (.93)	-.49 (.99)	
Total	.39 (.99)	-.18 (.93)	-.03 (1.02)	.04 (.97)	.05 (1.02)	-.16 (.98)	.37 (1.01)	-.14 (.94)	-.14 (1.00)	
Effect size	.06	.41	1.12	.33	.20	.77	.07	.47	.36	

Note. Overall mean across countries = 0, Standard deviation = 1.

Effect size estimates are Cohen's *d* and computed within country to compare High ADHD versus Low ADHD group.

Table 3
Mean, Standard Deviation and Loading for each DBRS Item by Time Point and Country

DBRS Items*	End of Preschool			End of 2nd grade		
	Australia (n = 253)	United States (n = 482)	Scandinavia (n = 239)	Australia (n = 95)	United States (n = 375)	Scandinavia (n = 128)
	<i>M (SD); Loading</i>	<i>M (SD); Loading</i>	<i>M (SD); Loading</i>	<i>M (SD); Loading</i>	<i>M (SD); Loading</i>	<i>M (SD); Loading</i>
IN Symptom						
1. Fails to attend	.65 (.63); 1.00	.65 (.66); 1.00	.46 (.61); 1.00	.86 (.66); 1.00	.72 (.65); 1.00	.58 (.56); 1.00
2. Trouble sustaining attention	.62 (.68); 1.25	.63 (.64); 1.05	.49 (.59); 0.95	.57 (.66); 0.99	.50 (.65); 1.16	.42 (.64); 1.84
3. Doesn't listen	.80 (.72); 1.20	.80 (.72); 1.28	.73 (.68); 0.95	.63 (.65); 0.70	.55 (.67); 1.12	.50 (.62); 1.75
4. Doesn't follow through	.87 (.75); 1.44	.79 (.72); 1.49	.41 (.54); 0.94	.79 (.65); 1.00	.62 (.70); 1.25	.34 (.55); 1.51
5. Difficulty organizing	.54 (.65); 1.22	.62 (.70); 1.30	.44 (.58); 0.98	.57 (.61); 0.99	.54 (.68); 1.26	.37 (.55); 1.34
6. Avoids tasks	.53 (.70); 1.35	.54 (.69); 1.07	.50 (.66); 0.99	.57 (.72); 0.94	.65 (.76); 1.13	.45 (.61); 1.30
7. Loses things	.57 (.67); 0.90	.55 (.67); 1.02	.45 (.56); 0.83	.52 (.63); 0.88	.55 (.69); 1.03	.34 (.49); 1.03
8. Easily distracted	.95 (.78); 1.55	.88 (.78); 1.54	.90 (.75); 0.87	.81 (.72); 1.04	.72 (.72); 1.33	.68 (.73); 1.76
9. Forgetful	.46 (.59); 1.08	.52 (.66); 1.16	.50 (.61); 0.68	.62 (.75); 1.13	.52 (.69); 1.21	.52 (.64); 1.53
H/I Symptom						
10. Fidgets	.81 (.87); 1.00	.82 (.87); 1.00	.76 (.80); 1.00	.63 (.72); 1.00	.53 (.72); 1.00	.53 (.76); 1.00
11. Leaves seat	1.05 (.85); 0.78	.59 (.76); 0.85	.74 (.73); 1.07	.62 (.69); 0.99	.28 (.54); 0.66	.37 (.58); 0.90
12. Runs or climbs	.63 (.78); 1.01	.61 (.78); 0.96	.45 (.65); 0.89	.33 (.55); 0.96	.26 (.59); 0.82	.16 (.43); 0.68
13. Difficulty playing quietly	.48 (.66); 0.79	.53 (.71); 0.94	.35 (.55); 0.76	.26 (.51); 1.05	.32 (.59); 0.95	.17 (.42); 0.60
14. On the go	.78 (.90); 1.27	.86 (.91); 1.11	.67 (.82); 1.17	.41 (.64); 1.06	.46 (.73); 1.09	.23 (.54); 0.80
15. Talks excessively	.95 (.89); 0.89	.90 (.88); .89	.55 (.71); 0.72	.61 (.89); 1.70	.57 (.74); 1.02	.30 (.58); 0.75
16. Blurts out answers	.67 (.75); 0.74	.63 (.72); 0.51	.30 (.57); 0.54	.57 (.77); 1.80	.51 (.66); 0.85	.40 (.58); 0.61
17. Difficulty waiting turn	1.04 (.76); 0.88	.89 (.74); 0.87	.90 (.71); 0.78	.57 (.65); 1.39	.49 (.67); 1.09	.46 (.64); 0.79
18. Interrupts	1.25 (.79); 0.83	1.16 (.80); 0.79	.71 (.66); 0.75	.80 (.72); 1.68	.70 (.72); 1.06	.36 (.62); 0.88

* *Note.* DBRS item content shortened for space. IN = inattention; H/I = hyperactivity/impulsivity.

Table 4

Model Fit Indices for Models of DSM – IV Attention-Deficit/Hyperactivity Disorder Symptoms

Model	χ^2	Dfs	CFI	RMSEA
End of Preschool ($n = 974$)				
1. One factor	1849.8 ***	135	.760	.114
2. IN and H/I factors	633.3 ***	129	.929	.063
3. Australia ($N = 253$)	302.8 ***	129	.908	.073
4. United States ($N = 482$)	432.1 ***	129	.920	.070
5. Scandinavia ($N = 239$)	267.7 ***	129	.916	.067
End of 2 nd Grade ($n = 598$)				
1. One factor	1535.4 ***	135	.729	.132
2. IN and H/I factors	561.0 ***	129	.917	.075
3. Australia ($N = 95$)	231.3 ***	129	.863	.092
4. United States ($N = 375$)	425.1 ***	129	.912	.078
5. Scandinavia ($N = 128$)	305.3 ***	129	.864	.104

Note. *DSM – IV* = Diagnostic and Statistical Manual of Mental Disorders (4th ed.); CFI = comparative fit index; RMSEA = root-mean-square error of approximation; *dfs* = degrees of freedom; IN = inattention; H/I = hyperactivity/impulsivity.

 $p < .001$.

Table 5

Model Fit Indices for Testing Weak Factorial Invariance

	All Countries Free	All Countries Constrained	USA and Australia Constrained	Australia and Scandinavia Constrained	USA and Scandinavia Constrained
End of Preschool ($n = 974$)					
χ^2	1002.70 ***	1072.70 ***	1027.81 ***	1036.51 ***	1049.16 ***
Dfs	387	419	403	403	403
CFI	.916	.911	.915	.913	.912
RMSEA	.040	.040	.040	.040	.041
χ^2		70.02	25.11	33.81	46.46
dfs		32	16	16	16
p value		< .001	.068	< .01	< .001
End of 2 nd Grade ($n = 598$)					
χ^2	962.90 ***	1033.43 ***	984.66 ***	997.68 ***	1009.77 ***
Dfs	387	419	403	403	403
CFI	.893	.886	.892	.890	.888
RMSEA	.050	.050	.049	.050	.050
χ^2		70.53	21.76	34.78	46.87
dfs		32	16	16	16
P value		< .001	.15	< .01	< .001

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; dfs = degrees of freedom.

 $p < .001$.

Table 6Comparison of Stability of Factor Loadings between End of Preschool and End of 2nd grade

	Free Model	All Factors Constrained	Inattention Constrained	Hyperactivity Constrained
Australia (<i>n</i> = 253)				
χ^2	557.06 ***	590.74 ***	569.96 ***	577.80 ***
<i>Dfs</i>	258	274	266	266
CFI	.887	.880	.885	.882
RMSEA	.058	.058	.057	.058
χ^2		33.685	12.901	20.741
<i>dfs</i>		16	8	8
<i>p</i> value		< .01	.115	< .01
United States (<i>n</i> = 482)				
χ^2	870.57 ***	917.44 ***	885.41 ***	902.54 ***
<i>Dfs</i>	258	274	266	266
CFI	.914	.910	.913	.911
RMSEA	.053	.052	.052	.053
χ^2		46.862	14.839	31.966
<i>dfs</i>		16	8	8
<i>p</i> value		< .001	.062	< .001
Scandinavia (<i>n</i> = 239)				
χ^2	600.69 ***	639.522 ***	627.70 ***	612.60 ***
<i>Dfs</i>	258	274	266	266
CFI	.883	.876	.877	.882
RMSEA	.060	.060	.061	.060
χ^2		38.832	27.01	11.91
<i>dfs</i>		16	8	8
<i>p</i> value		< .01	< .001	.155

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; *dfs* = degrees of freedom.

p < .001.